Scaling Frog Monitoring with FrogID: A Robust Classification Pipeline for Citizen Science Using Bioacoustic Foundation Models

Julia Tan
FrogID
Australian Museum Research Institute
Sydney, NSW, Australia
Julia.Tan@Australian.Museum

Jodi J. L. Rowley [*]
FrogID
Australian Museum Research Institute
Sydney, NSW, Australia
Jodi.Rowley@Australian.Museum

Abstract

Amidst global biodiversity declines, audio-based citizen science projects offer significant potential for biodiversity monitoring, but the need for manual validation limits scalability. The FrogID project has gathered over 1.3 million frog records from over 800,000 audio submissions, advancing amphibian research and conservation in Australia. However, manual species identification remains time-consuming, creating backlogs, delaying conservation action, and reducing user engagement. We present a frog species identification pipeline that combines unsupervised source separation with an audio-language foundation model to refine coarse annotations, followed by transfer learning from cross-taxa embeddings with a hybrid classifier. The method achieves strong per-species performance even on non-quality audio, enabling scalable frog monitoring.

1 Introduction/Motivation

In the midst of a global biodiversity crisis, there is an urgent need to collect species occurrence data to monitor trends and evaluate management actions [1]. Community science has emerged as a powerful tool to address this challenge, with audio-based approaches offering unique opportunities to monitor acoustically active taxa at unprecedented scales. Amphibians are an ideal focal taxon, as they are the most threatened vertebrate group, perform vital ecosystem functions, serve as sensitive indicators of environmental health, and largely rely on acoustic communication. The Australian FrogID project [2] exemplifies this potential: through a free smartphone app, it has engaged over 100,000 participants to generate more than 1.3 million frog occurrence records from over 800,000 geo-referenced recordings in under eight years. FrogID has advanced our understanding of the impacts of landuse modification and extreme weather events [3–8], is a vital tool for invasive species detection [9, 10], and has helped discover 13 species new to science [11–14].

Currently, human experts listen to each 20-60s recording and label it with one or more frog species names, or as “not-a-frog” (i.e. birds, insects or other noise). This reliance on manual validation leads to a substantial backlog when submission rates are high, causing delays of weeks to months, which can result in user dissatisfaction, as well as slow detection and reporting of invasive species outside of their known range. Advances in AI could help to expedite the validation of priority recordings. While citizen science is capable of rapid data collection across large spatial scales, existing AI methods that leverage such data present the following main challenges:

[*] Also affiliated with the Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW, Australia. Alternative email: jodi.rowley@unsw.edu.au.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI for Non-Human Animal Communication.

- **Poor representation of non-bird animal taxa** - Most bioacoustic datasets are dominated by bird recordings, causing models to bias toward bird taxa, and underperform on amphibians, insects and mammals [15].
- **Noisy and/or coarse labels in citizen science datasets** - Due to the participatory nature of citizen science, these datasets can include misidentifications and clip-level annotations, weakening supervision and reducing training accuracy [16].
- **Disparity between curated benchmark datasets and the "real-world"** - Commonly-used benchmark datasets typically contain only focal recordings, resulting in a performance gap when models are applied in real-world settings and faced with soundscape complexity [17].

To address the challenges of citizen science data, we propose a robust method for converting coarse-labelled bioacoustic data into an efficient training dataset for a frog species classifier model. Importantly, our pipeline employs cross-taxa pre-trained models to enable rapid adaptation from limited training examples, with generalisability to future studies across other animal groups. We will demonstrate that our method is capable of dealing with real-world audio conditions despite minimal supervision, making it well-suited for use in citizen science applications. Ultimately, this will allow large-scale citizen science data to be leveraged for greater ecological impact and effective species monitoring.

2 Methodology

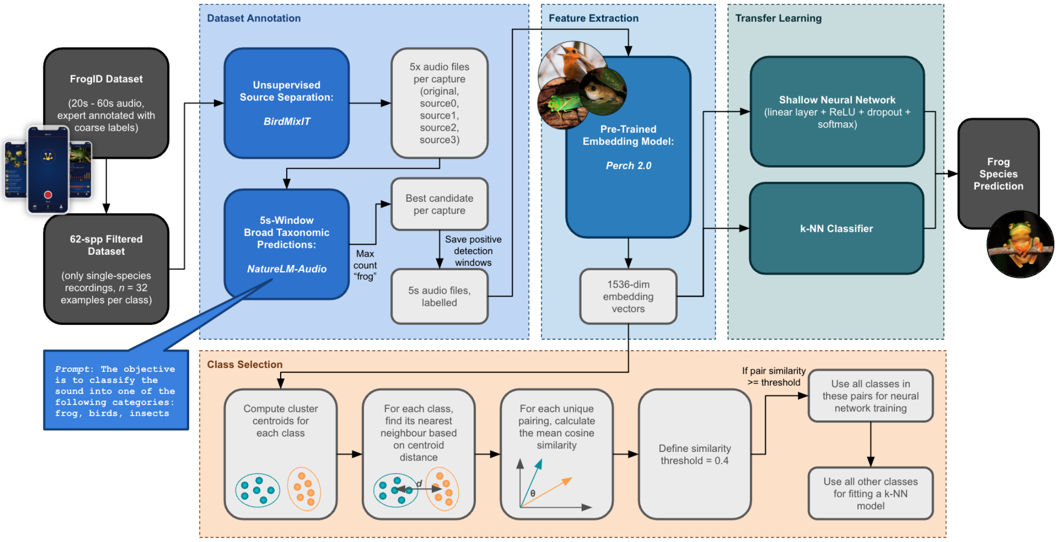

Dataset Annotation
Feature Extraction
Transfer Learning
FrogID Dataset
(20s - 60s audio, expert annotated with coarse labels)
62-spp Filtered Dataset
(only single-species recordings, n = 32 examples per class)
Unsupervised Source Separation:
BirdMixIT
5x audio files per capture (original, source0, source1, source2, source3)
5s-Window Broad Taxonomic Predictions:
NatureLM-Audio
Max count "frog"
Best candidate per capture
Save positive detection windows
5s audio files, labelled
Prompt: The objective is to classify the sound into one of the following categories: frog, birds, insects
Pre-Trained Embedding Model:
Perch 2.0
1536-dim embedding vectors
Shallow Neural Network
(linear layer + ReLU + dropout + softmax)
k-NN Classifier
Frog Species Prediction
Class Selection
Compute cluster centroids for each class
For each class, find its nearest neighbour based on centroid distance
d
For each unique pairing, calculate the mean cosine similarity
θ
Define similarity threshold = 0.4
If pair similarity >= threshold
Use all classes in these pairs for neural network training
Use all other classes for fitting a k-NN model

Figure 1: Overall pipeline for using coarse-labelled FrogID audio data to train a frog species classifier, involving a hybrid neural network and k-NN approach.

2.1 Dataset Filtering

Starting from nearly 800,000 captures across 231 Australian frog species, a general quality and content filter is applied to reduce label noise. Then, all multi-species calls are removed, to eliminate ambiguity around label assignment. To prevent overfitting to rare taxa, a minimum threshold of 32 captures per species is implemented, ensuring sufficient examples. This yields a final dataset of 13,026 calls across 62 species. For the exact class breakdown (prior to class balancing), see subsection A.1.

Table 1: Filtering stages for converting the raw FrogID dataset into a smaller training dataset of audio captures.

Filtering Stage	Total Dataset Size (# of Unique Captures)	# of Unique Classes
All frog captures (as of checkpoint March 2025)	798947	231
Quality call, no people activity or inappropriate content, not a duplicate	19120	231
Single-species call	14096	185
Species contains at least 32 captures	13026	62

2.2 Dataset Annotation

To extract the target frog segments from the FrogID recordings, while filtering out non-target sounds such as bird or insect vocalisations, we developed a novel annotation pipeline (Figure 1) that applies BirdMixIT (Apache-2.0) [18] for unsupervised source separation to isolate candidate vocal streams, followed by NatureLM-Audio (CC-BY-NC-SA-4.0) [19] for broad taxonomic classification as either *frog*, *bird* or *insect* in 5s-windows. The best candidate audio file is selected as the one with the highest proportion of frog-positive windows, ensuring optimal alignment with the single-species label. All frog-positive windows from this file are then saved as separate audio files.

2.3 Transfer Learning Approach

To extract reliable features from the 5s training examples, we utilise Perch 2.0 (Apache-2.0), a bioacoustic embedding model trained on over one million recordings from diverse taxa and environments [20]. This comprehensive corpus allowed Perch 2.0 to learn generalisable acoustic features that transfer effectively across species, including those not represented in the pre-training data. To qualitatively assess its applicability to the target frog species, we applied T-SNE to project the 1536-dimensional embedding vectors into a 2D visualisation (Figure 2).

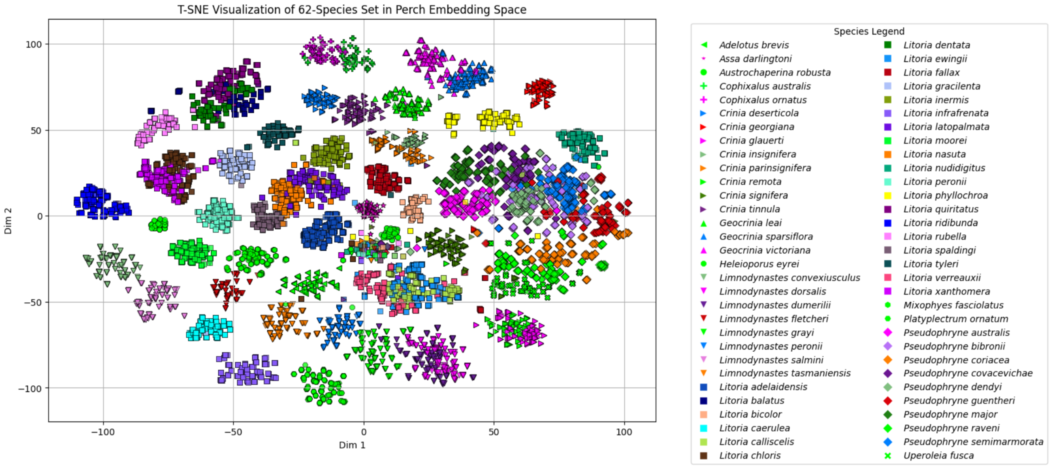


Figure 2: T-SNE 2D visualisation of 1536-dimensional embedding vectors obtained through Perch 2.0, representing 62 frog species from the FrogID audio dataset, with marker shape denoting genus and marker colour denoting species.

2.4 Class Selection Strategy

We adopt a hybrid strategy that categorises species according to their acoustic distinctiveness in the Perch 2.0 embedding space. Classes that form clear, well-separated clusters are considered easy to distinguish, and therefore suitable for a simple k-nearest neighbour (k-NN) approach. In contrast, classes that show high inter-cluster similarity to their nearest neighbour are designated for neural network training, whereby non-linear decision boundaries can be learned. Using a cosine similarity threshold of 0.4, 29 out of the total 62 species are assigned for neural network training (see subsection A.2). This class selection strategy reduces unnecessary model complexity, while allocating greater capacity to challenging cases.

2.5 Experiment Setup

The shallow neural network model was trained for a total of 50 epochs, with the best model kept based on minimising validation loss. For all other training parameters, see subsection A.2.

2.6 Ethical Considerations

Several ethical safeguards were incorporated into the system design to ensure responsible handling of sensitive biodiversity and participant data:

- There is no geolocation metadata or Personally Identifiable Information (PII) anywhere in the dataset
- The minimum class size filter removes small populations from the dataset
- The no people activity or inappropriate content filter removes any potentially identifiable audio from the dataset

3 Results

Table 2: Performance summary metrics for frog species classifier, showing weighted average scores over a total of 62 Australian frog species.

	Validation Set				**Holdout Set**				**Non-Quality Audio**			
	Precision	Recall	F1	Support	Precision	Recall	F1	Support	Precision	Recall	F1	Support
Neural Net	0.8975	0.8971	0.8956	3916	0.8455	0.7968	0.8099	14890	0.9378	0.7889	0.8449	379
k-NN (k=7)	0.9638	0.9620	0.9618	1288	-	-	-	-	0.9715	0.9516	0.9603	1673

3.1 Species Classifier Performance

The neural network classifier achieved consistently high performance across the selected 29 species, with precision, recall and F1-scores typically > 0.85. *Pseudophryne bibronii* was the notable exception (Figure 3), due to misclassifications with congeneric species, primarily *Pseudophryne dendyi* (see subsection A.3). The k-NN classifier performed very strongly across all other species, with the average precision, recall and F1 scores all > 0.95.

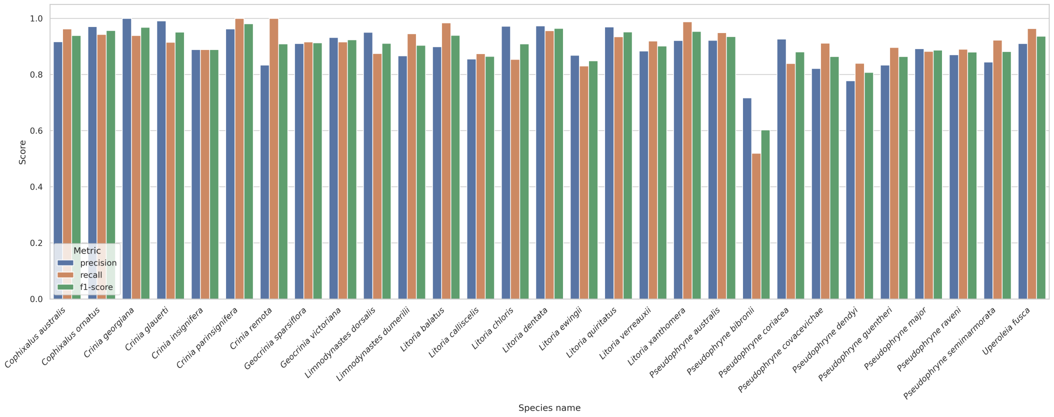


Figure 3: Per-class performance of the neural network classifier across 29 Australian frog species from the FrogID dataset, evaluated on the validation set.

3.2 Evaluation on Non-Quality Audio

To assess robustness under varying audio conditions, we evaluated both classifiers on FrogID recordings that were not tagged by human experts as “high-quality” audio (see subsection A.1 for an example). Overall performance for the neural network classifier declined relative to the validation set, due to a drop in recall. On the other hand, the k-NN classifier maintains equivalent performance.

3.3 Ablation Study

To quantify the effects of each pipeline component on overall model performance, we conducted an ablation study for the neural network classifier. The proposed pipeline was compared against two alternate configurations: one with only source separation disabled and the other with the full dataset annotation step removed. The evaluation results in Figure 4 show that performance degradation across all metrics is most pronounced on the non-quality audio recordings, confirming that the proposed pipeline contributes most to robustness under challenging real-world acoustic conditions.

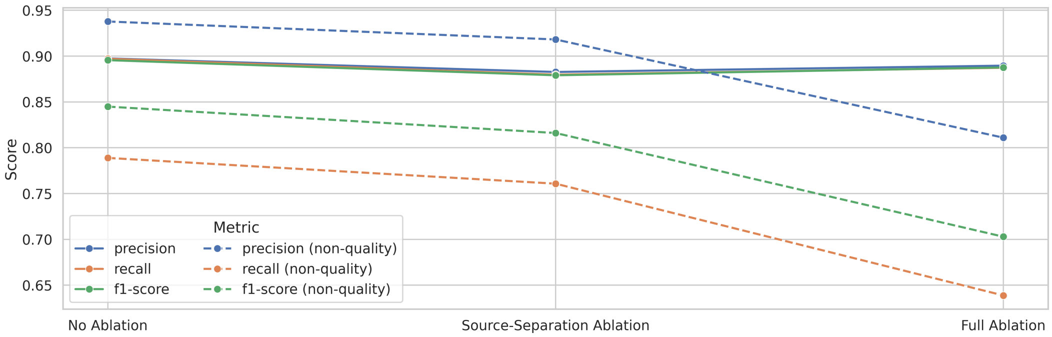


Figure 4: Performance summary metrics for the neural network classifier under two ablation conditions: removal of source separation only, and removal of dataset annotation entirely (i.e. the original audio file is just segmented into 5s windows). Results on the validation set (solid lines) show a modest decline in performance, while those on the non-quality audio recordings (dashed lines) exhibit the most significant reductions across all metrics.

4 Conclusion

This work presents a novel frog species identification framework, demonstrating how coarsely labelled citizen science data can be transformed into effective training datasets. We employ cross-taxa pretrained models and a hybrid model approach, achieving high accuracy with minimal training examples required. The approach performs robustly even on recordings with real-world complexity, enabling scalable species monitoring and more timely management actions. Thus, by coupling advances in bioacoustic AI with citizen science, we provide a foundation for facilitating large-scale audio data collection aimed at supporting informed conservation decisions to halt biodiversity loss.

5 Acknowledgements

We would like to thank the generous support of Holly Corn and Jonathan Kaufelt and the Innovative Biodiversity Monitoring Grant from the Australian Federal Department of Climate Change, Energy, the Environment and Water. We also thank the Citizen Science Grants of the Australian Government and the Impact Grants program of IBM Australia for providing funding and resources to help build the initial FrogID app; the generous donors who have provided funding for the project; the Museum and Art Gallery of the Northern Territory, Queensland Museum, South Australian Museum, Tasmanian Museum and Art Gallery, and Western Australian Museum as FrogID partner museums; the many Australian Museum staff and volunteers who make up the FrogID team; and, most importantly, the thousands of citizen scientists across Australia who have volunteered their time to record frogs.

References

[1] V. Proença *et al.*, "Global biodiversity monitoring: From data sources to essential biodiversity variables," *Biological Conservation*, vol. 213, pp. 256–263, 2017, SI:Measures of biodiversity, ISSN: 0006-3207. DOI: `https://doi.org/10.1016/j.biocon.2016.07.014`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0006320716302786`.

[2] J. Rowley *et al.*, "Frogid: Citizen scientists provide validated biodiversity data on frogs of australia," *Herpetological Conservation and Biology*, vol. 14, pp. 155–170, Apr. 2019.

[3] C. Callaghan, G. Liu, B. Mitchell, A. Poore, and J. Rowley, "Urbanization negatively impacts frog diversity at continental, regional, and local scales," *Basic and Applied Ecology*, vol. 54, Apr. 2021. DOI: `10.1016/j.baae.2021.04.003`.

[4] G. Liu, J. Rowley, R. Kingsford, and C. Callaghan, *Species' traits drive amphibian tolerance to anthropogenic habitat modification*, May 2021. DOI: `10.32942/OSF.IO/KU4MV`.

[5] B. Mitchell, C. Callaghan, and J. Rowley, "Continental-scale citizen science data reveal no changes in acoustic responses of a widespread tree frog to an urbanisation gradient," *Journal of Urban Ecology*, vol. 6, Feb. 2020. DOI: `10.1093/jue/juaa002`.

[6] M. E. Malerba *et al.*, "Conserving nature's chorus: Local and landscape features promoting frog species richness in farm dams," *Biological Conservation*, vol. 286, p. 110 270, 2023, ISSN: 0006-3207. DOI: `https://doi.org/10.1016/j.biocon.2023.110270`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0006320723003713`.

[7] B. A. Mitchell, S. B. Z. Gorta, C. T. Callaghan, R. T. Kingsford, and J. J. L. Rowley, "Fighting the flames: site-specific effects determine species richness of Australian frogs after fire," *Wildlife Research*, vol. 51, no. 1, 2023. DOI: `10.1071/WR22175`. [Online]. Available: `https://doi.org/10.1071/WR22175`.

[8] J. Rowley, C. Callaghan, and W. Cornwell, "Widespread short-term persistence of frog species after the 2019–2020 bushfires in eastern australia revealed by citizen science," *Conservation Science and Practice*, vol. 2, Sep. 2020. DOI: `10.1111/csp2.287`.

[9] T. Cutajar, C. Portway, G. Gillard, and J. Rowley, "Australian frog atlas: Species' distribution maps informed by the frogid dataset," *Technical Reports of the Australian Museum online*, vol. 36, pp. 1–48, Jun. 2022. DOI: `10.3853/j.1835-4211.36.2022.1789`.

[10] J. J. L. Rowley and C. T. Callaghan, "Tracking the spread of the eastern dwarf tree frog (Litoria fallax) in Australia using citizen science," *Australian Journal of Zoology*, vol. 70, no. 6, pp. 204–210, 2023. DOI: `10.1071/ZO23012`. [Online]. Available: `https://doi.org/10.1071/ZO23012`.

[11] T. PARKIN *et al.*, "Systematic assessment of the brown tree frog (anura: Pelodryadidae: Litoria ewingii) reveals two endemic species in south australia," *Zootaxa*, vol. 5406, pp. 1–36, Feb. 2024. DOI: `10.11646/zootaxa.5406.1.1`.

[12] T. Parkin, J. Rowley, G. Gillard, J. Sopniewski, G. Shea, and S. Donnellan, "Systematics and taxonomy of the northern banjo frog (anura: Limnodynastidae: Limnodynastes terraereginae) and allied taxa," *Ichthyology Herpetology*, vol. 112, Mar. 2024. DOI: `10.1643/h2023025`.

[13] W. PURSER *et al.*, "Systematics of the little red tree frog, litoria rubella (anura: Pelodryadidae), with the description of two new species from eastern australia and arid western australia," *Zootaxa*, vol. 5594, pp. 269–315, Feb. 2025. DOI: `10.11646/zootaxa.5594.2.3`.

[14] J. Rowley *et al.*, "Two new frog species from the litoria rubella species group from eastern australia," *Zootaxa*, vol. 5071, pp. 1–41, Nov. 2021. DOI: 10.11646/zootaxa.5071.1.1.

[15] B. van Merriënboer, J. Hamer, V. Dumoulin, E. Triantafillou, and T. Denton, "Birds, bats and beyond: Evaluating generalization in bioacoustics models," *Frontiers in Bird Science*, vol. Volume 3 - 2024, 2024, ISSN: 2813-3870. DOI: 10.3389/fbirs.2024.1369756. [Online]. Available: https://www.frontiersin.org/journals/bird-science/articles/10.3389/fbirs.2024.1369756.

[16] F. Michaud, J. Sueur, M. Le Cesne, and S. Haupert, "Unsupervised classification to improve the quality of a bird song recording dataset," *Ecological Informatics*, vol. 74, p. 101 952, 2023, ISSN: 1574-9541. DOI: https://doi.org/10.1016/j.ecoinf.2022.101952. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574954122004022.

[17] M. Boudiaf, T. Denton, B. van Merriënboer, V. Dumoulin, and E. Triantafillou, *In search for a generalizable method for source free domain adaptation*, 2023. arXiv: 2302.06658 [cs.LG]. [Online]. Available: https://arxiv.org/abs/2302.06658.

[18] T. Denton, S. Wisdom, and J. R. Hershey, *Improving bird classification with unsupervised sound separation*, 2021. arXiv: 2110.03209 [eess.AS]. [Online]. Available: https://arxiv.org/abs/2110.03209.

[19] D. Robinson *et al.*, *Naturelm-audio: An audio-language foundation model for bioacoustics*, 2025. arXiv: 2411.07186 [cs.SD]. [Online]. Available: https://arxiv.org/abs/2411.07186.

[20] B. Merriënboer, V. Dumoulin, J. Hamer, L. Harrell, A. Burns, and T. Denton, *Perch 2.0: The bittern lesson for bioacoustics*, Aug. 2025. DOI: 10.48550/arXiv.2508.04665.

A Technical Appendices and Supplementary Material

A.1 FrogID Dataset Analysis

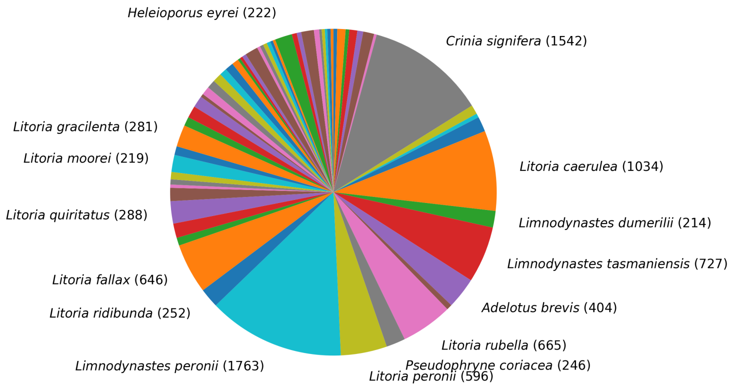
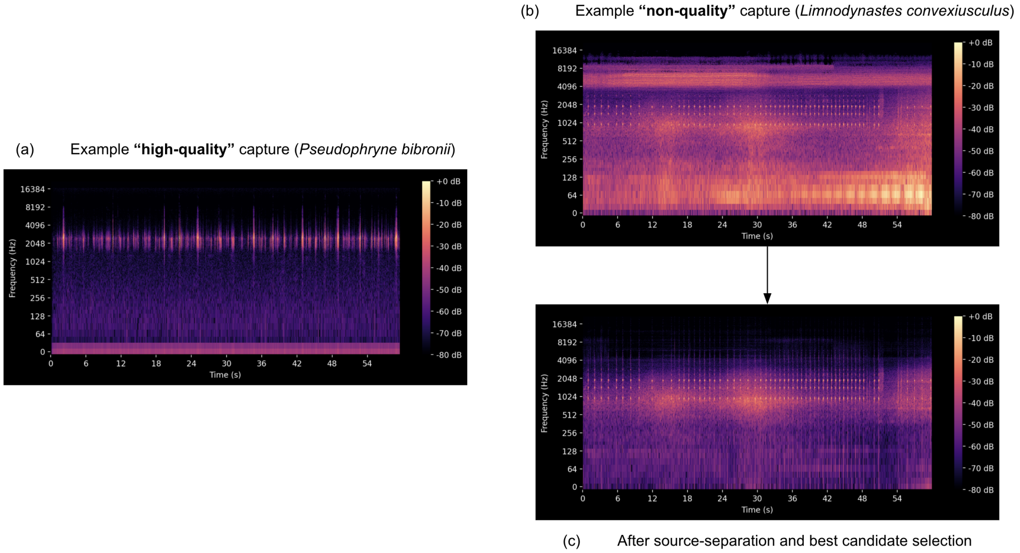


Figure 5: Class distribution after applying filtering stages to the FrogID dataset, with the largest 15 classes labelled using the scientific name of the species and corresponding number of captures.


(a) Example **"high-quality"** capture (*Pseudophryne bibronii*)

(b) Example **"non-quality"** capture (*Limnodynastes convexiusculus*)

(c) After source-separation and best candidate selection

Figure 6: Comparison between an example FrogID submission that was tagged as "high-quality" (left) versus "non-quality" (right). The "non-quality" capture is also shown after source-separation and best candidate selection has been applied (bottom).

A.2 Training Data and Parameter Selection

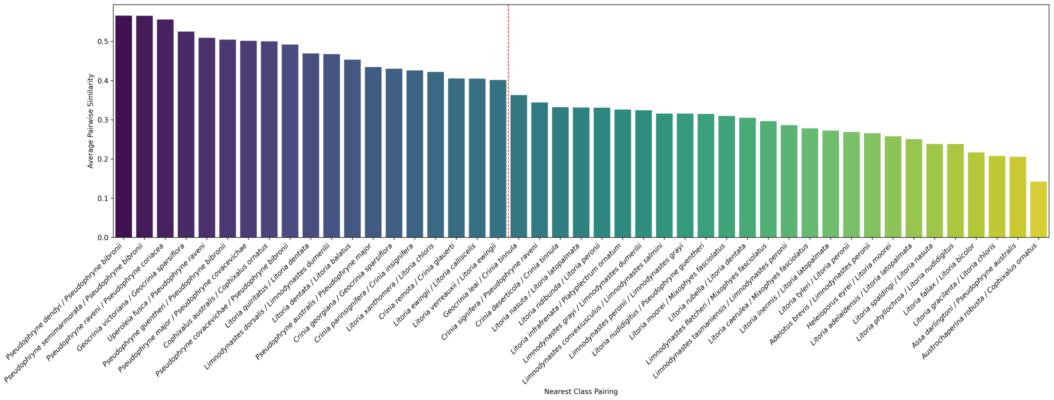


Figure 7: Inter-cluster cosine similarity scores for all nearest class pairs in the filtered FrogID audio dataset, calculated using the 1536-dimensional embedding vectors generated with Perch 2.0. The chosen threshold = 0.4 is shown (red), with all 29 species on the left of the threshold being selected for neural network classifier training.

Table 3: Training parameters used for neural network classifier, which was trained on top of frozen Perch 2.0 embeddings for 29 Australian frog species from the FrogID audio dataset.

Training Parameter	**Value**
Batch size	32
Train-test split	80-20
Stratified split	True
Class balancing (n = 32)	True
Optimiser	Adam
Loss function	Categorical cross-entropy
Epochs	50
Base learning rate	1e-4
Learning rate warmup steps	255
Cosine annealing first decay steps	850

A.3 Per-Class Performance of Neural Network Classifier

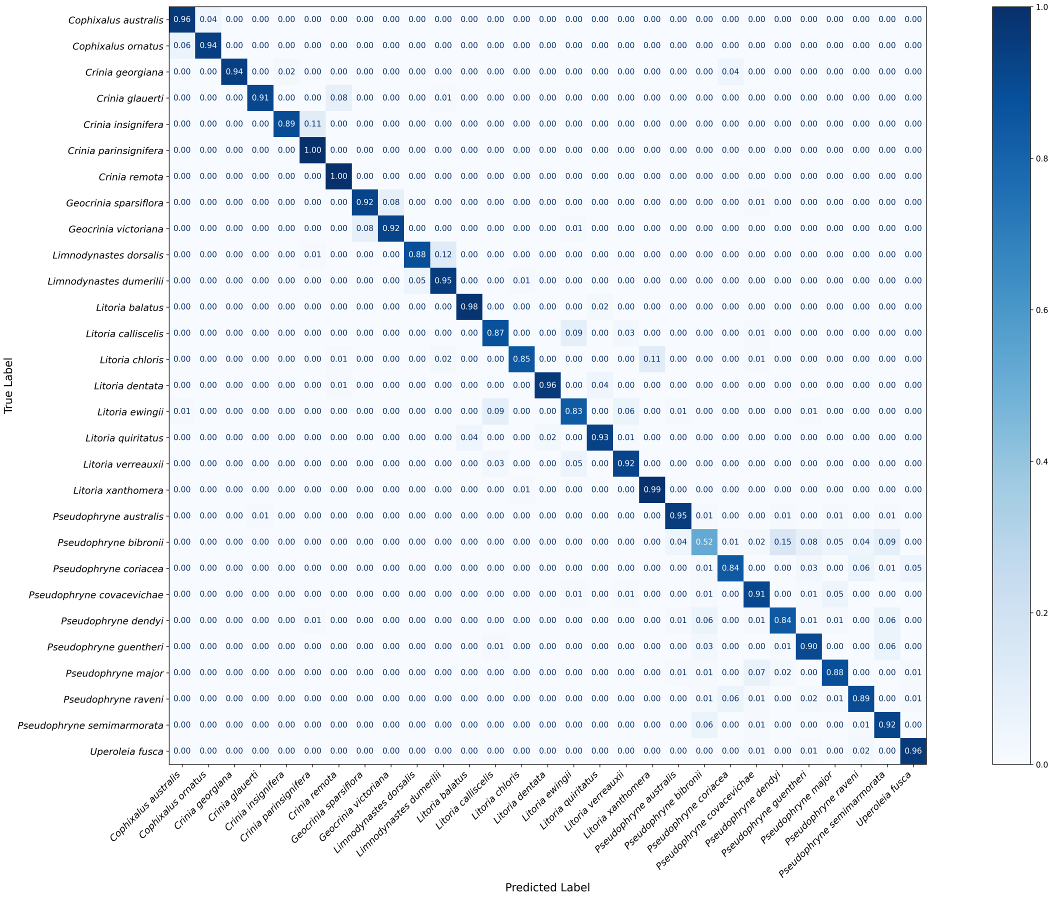


Figure 8: Normalised confusion matrix showing per-class classification performance of the neural network trained on FrogID audio captures of 29 Australian frog species, evaluated on the validation set.

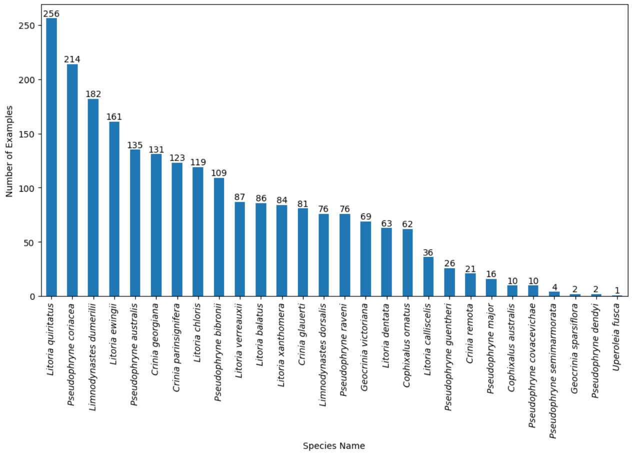


Figure 9: Class distribution of the holdout set, which is the remaining captures from the filtered FrogID dataset that were not randomly selected for training or validation of the neural network classifier, across the same 29 species. *Crinia insignifera* is not included since all captures were already used for either training or validation.

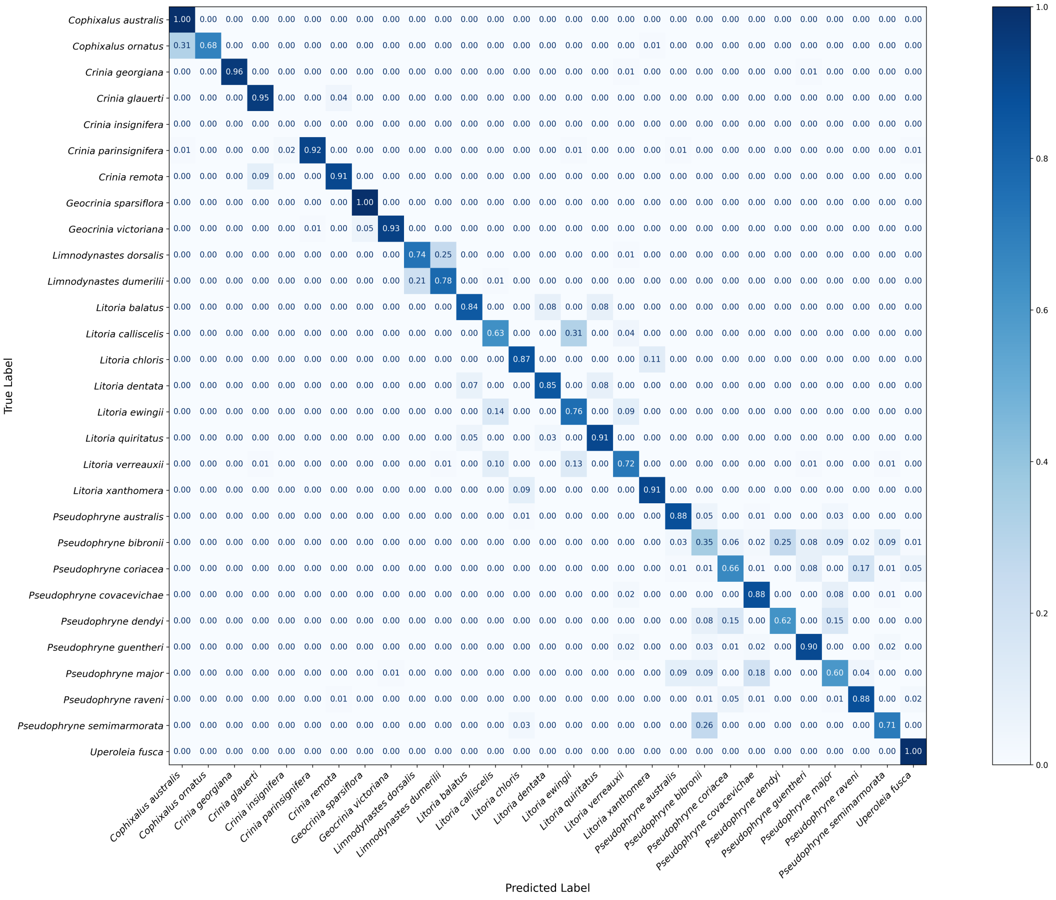
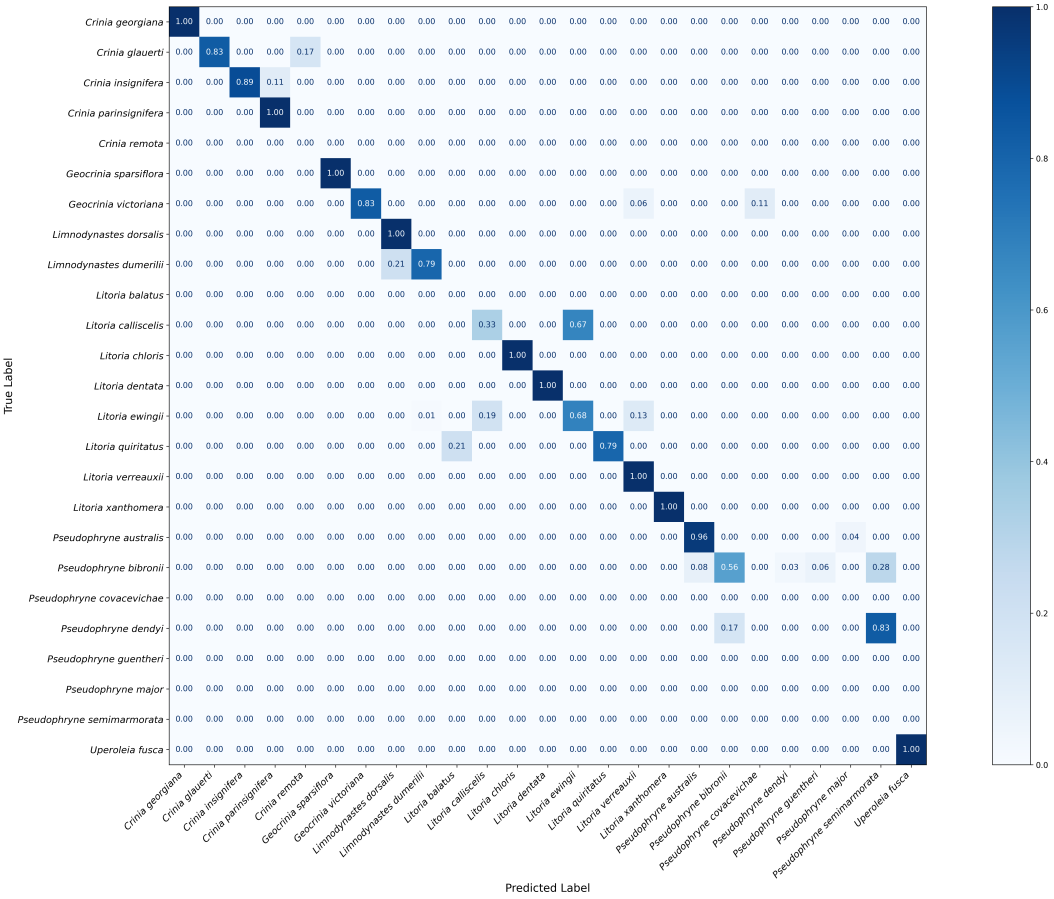


Figure 10: Normalised confusion matrix showing per-class classification performance of the neural network trained on FrogID audio captures of 29 Australian frog species, evaluated on a larger holdout set from the FrogID dataset.



	Crinia georgiana	Crinia glauerti	Crinia insignifera	Crinia parinsignifera	Crinia remota	Geocrinia sparsiflora	Geocrinia victoriana	Limnodynastes dorsalis	Limnodynastes dumerilii	Litoria balatus	Litoria calliscelis	Litoria chloris	Litoria dentata	Litoria ewingii	Litoria quiritatus	Litoria verreauxii	Litoria xanthomera	Pseudophryne australis	Pseudophryne bibronii	Pseudophryne covacevichae	Pseudophryne dendyi	Pseudophryne guentheri	Pseudophryne major	Pseudophryne semimarmorata	Uperoleia fusca
Crinia georgiana	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Crinia glauerti	0.00	0.83	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Crinia insignifera	0.00	0.00	0.89	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Crinia parinsignifera	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Crinia remota	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Geocrinia sparsiflora	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Geocrinia victoriana	0.00	0.00	0.00	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00
Limnodynastes dorsalis	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Limnodynastes dumerilii	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Litoria balatus	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Litoria calliscelis	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Litoria chloris	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Litoria dentata	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Litoria ewingii	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.19	0.00	0.00	0.68	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Litoria quiritatus	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.00	0.00	0.00	0.00	0.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Litoria verreauxii	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Litoria xanthomera	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pseudophryne australis	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.00	0.00	0.00	0.00	0.04	0.00	0.00
Pseudophryne bibronii	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.56	0.00	0.03	0.06	0.00	0.28	0.00
Pseudophryne covacevichae	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pseudophryne dendyi	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.83	0.00
Pseudophryne guentheri	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pseudophryne major	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pseudophryne semimarmorata	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Uperoleia fusca	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Figure 11: Normalised confusion matrix showing per-class classification performance of the neural network trained on FrogID audio captures of 29 Australian frog species, evaluated on low-quality audio examples from the FrogID dataset.

A.4 K-Fold Cross Validation for k-NN Classifier

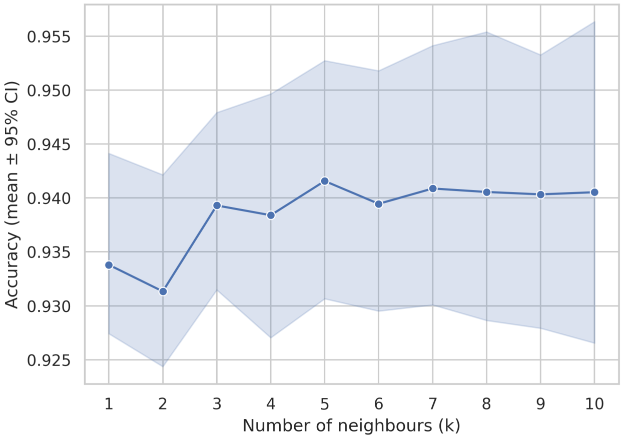


Figure 12: Classification accuracy across varying neighbourhood sizes for the k-NN model fitted on 33 Australian frog species, showing the mean score over 5 folds with ± 95% confidence interval (shaded).

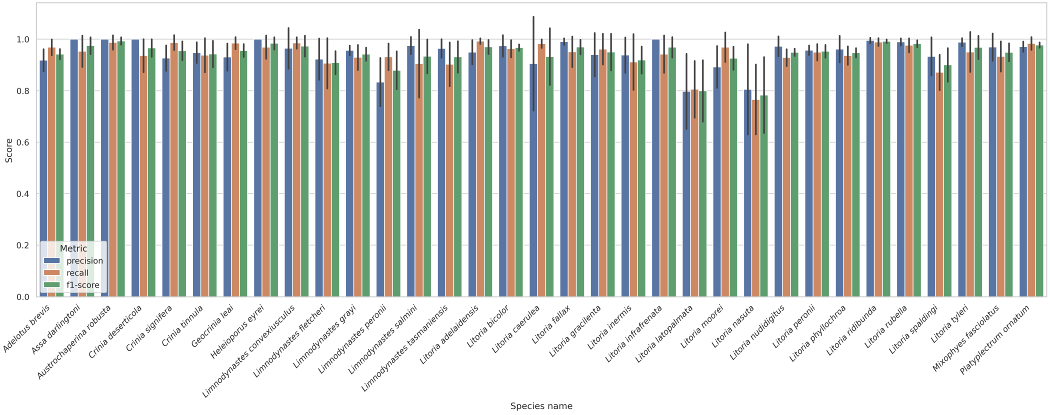


Figure 13: Per-class performance of the k-NN model (using k=7) across 33 Australian frog species, showing mean metrics over 5 folds with ±1 standard deviation error bars.

A.5 Compute Resources Used

Table 4: Compute resources used in the experiments, including all pretrained models as well as the full embedding dataset of 62 frog species, obtained from running Perch 2.0 on the FrogID audio dataset.

Resource Name	**Value**
Worker type	1x GPU
Provider type	Internal cluster
Memory used for loading NatureLM-audio	15799 MB
Total memory used for loading all models in the pipeline	16295 MB
Total memory used for storing the embedding dataset	74.43 MB
Compute time for generating the full embedding dataset of 62 species	2.5 minutes
Compute time for a single training run of the neural network classifier	77 minutes